

# GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND METHODS OF USE THEREOF

Line B1

## BACKGROUND OF THE INVENTION

### Field of the Invention

The present invention describes nucleic acid sequences for eukaryotic genes encoding  
5  $\epsilon$  lycopene  $\epsilon$ -cyclase (also known as  $\epsilon$ -cyclase and  $\epsilon$  lycopene cyclase), isopentenyl  
pyrophosphate isomerase (IPP) and  $\beta$ -carotene hydroxylase as well as vectors containing the  
same and hosts transformed with said vectors. The present invention also provides methods  
10 for augmenting the accumulation of carotenoids, changing the composition of the  
carotenoids, and producing novel and rare carotenoids. The present invention provides  
methods for controlling the ratio or relative amounts of various carotenoids in a host. The  
invention also relates to modified lycopene  $\epsilon$ -cyclase, IPP isomerase and  $\beta$ -carotene  
15 hydroxylase. Additionally, the present invention provides a method for screening for genes  
and cDNAs encoding enzymes of carotenoid biosynthesis and metabolism.

## Background of the Invention

Carotenoid pigments with cyclic endgroups are essential components of the photosynthetic apparatus in oxygenic photosynthetic organisms (e.g., cyanobacteria, algae and plants; Goodwin, 1980). The symmetrical bicyclic yellow carotenoid pigment  $\beta$ -carotene (or, in rare cases, the asymmetrical bicyclic  $\alpha$ -carotene) is intimately associated with the photosynthetic reaction centers and plays a vital role in protecting against potentially lethal photooxidative damage (Koyama, 1991).  $\beta$ -carotene and other carotenoids derived from it or from  $\alpha$ -carotene also serve as light-harvesting pigments (Siefermann-Harms, 1987), are involved in the thermal dissipation of excess light energy captured by the light-harvesting antenna (Demmig-Adams & Adams, 1992), provide substrate for the biosynthesis of the plant growth regulator abscisic acid (Rock & Zeevaart, 1991; Parry & Horgan, 1991), and are precursors of vitamin A in human and animal diets (Krinsky, 1987). Plants also exploit carotenoids as coloring agents in flowers and fruits to attract pollinators and agents of seed dispersal (Goodwin, 1980). The color provided by carotenoids is also of agronomic value in a number of important crops. Carotenoids are currently harvested from a variety of organisms, including plants, algae, yeasts, cyanobacteria and bacteria, for use as pigments in food and feed.

The probable pathway for formation of cyclic carotenoids in plants, algae and cyanobacteria is illustrated in Figure 1. Two types of cyclic endgroups or rings are commonly found in higher plant carotenoids, these are referred to as the  $\beta$  (*beta*) and  $\epsilon$  (*epsilon*) rings (Fig. 3). The precursor acyclic endgroup (no ring structure) is referred to as the  $\Psi$  (*psi*) endgroup. The  $\beta$  and  $\epsilon$  endgroups differ only in the position of the double bond in the ring. Carotenoids with two  $\beta$  rings are ubiquitous, and those with one  $\beta$  and one  $\epsilon$  ring are common, but carotenoids with two  $\epsilon$  rings are uncommon.  $\beta$ -carotene (Fig. 1) has two  $\beta$ -endgroups and is a symmetrical compound that is the precursor of a number of other important plant carotenoids such as zeaxanthin and violaxanthin (Fig. 2).

Genes encoding enzymes of carotenoid biosynthesis have previously been isolated from a variety of sources including bacteria (Armstrong et al., 1989, Mol. Gen. Genet. 216, 254-268; Misawa et al., 1990, J. Bacteriol., 172, 6704-12), fungi (Schmidhauser et al., 1990, Mol. Cell. Biol. 10, 5064-70), cyanobacteria (Chamovitz et al., 1990, Z. Naturforsch. 45c, 482-86; Cunningham et al., 1994) and higher plants (Bartley et al., Proc. Natl. Acad. Sci USA 88, 6532-36; Martinez-Ferez & Vioque, 1992, Plant Mol. Biol. 18, 981-83). Many of the isolated enzymes show a great diversity in structure, function and inhibitory properties between sources. For example, phytoene desaturases from the cyanobacterium *Synechococcus* and from higher plants and green algae carry out a two-step desaturation to yield  $\zeta$ -carotene as a reaction product. In plants and cyanobacteria a second enzyme ( $\zeta$ -carotene desaturase), similar in amino acid sequence to the phytoene desaturase, catalyzes two additional desaturations to yield lycopene. In contrast, a single desaturase enzyme from *Erwinia herbicola* and from other bacteria introduces all four double bonds required to form lycopene. The *Erwinia* and other bacterial desaturases bear little amino acid sequence similarity to the plant and cyanobacterial desaturase enzymes, and are thought to be of unrelated ancestry. Therefore, even with a gene in hand from one source, it may be difficult to identify a gene encoding an enzyme of similar function in another organism. In particular, the sequence similarity between certain of the prokaryotic and eukaryotic genes encoding enzymes of carotenoid biosynthesis is quite low.

Further, the mechanism of gene expression in prokaryotes and eukaryotes appears to differ sufficiently such that one cannot expect that an isolated eukaryotic gene will be properly expressed in a prokaryotic host.

The difficulties in isolating genes encoding enzymes with similar functions is exemplified by recent efforts to isolate the gene encoding the enzyme that catalyzes the formation of  $\beta$ -carotene from the acyclic precursor lycopene. Although a gene encoding an enzyme with this function had been isolated from a bacterium, it had not been isolated from any photosynthetic prokaryote or from any eukaryotic organism. The isolation and characterization of the enzyme catalyzing formation of  $\beta$ -carotene in the cyanobacterium *Synechococcus* PCC7942 was described by the present inventors and others (Cunningham et al., 1993 and 1994). The amino acid sequence similarity of the cyanobacterial enzyme to the various bacterial lycopene  $\beta$ -cyclases is so low (ca. 18-25% overall; Cunningham et al., 1994) that there is much uncertainty as to whether they share a common ancestry or, instead, represent an example of convergent evolution.

The need remains for the isolation of eukaryotic and prokaryotic genes and cDNAs encoding polypeptides involved in the carotenoid biosynthetic pathway, including those encoding a lycopene  $\epsilon$ -cyclase, IPP isomerase and  $\beta$ -carotene hydroxylase. There remains a need for methods to enhance the production of carotenoids, to alter the composition of carotenoids, and to reduce or eliminate carotenoid production. There also remains a need in the art for methods for screening for genes and cDNAs encoding enzymes of carotenoid biosynthesis and metabolism.

#### SUMMARY OF THE INVENTION

Accordingly, a first object of this invention is to provide purified and/or isolated nucleic acids which encode enzymes involved in carotenoid biosynthesis; in particular, lycopene  $\epsilon$ -cyclase, IPP isomerase and  $\beta$ -carotene hydroxylase.

A second object of this invention is to provide purified and/or isolated nucleic acids which encode enzymes which produce novel or uncommon carotenoids.

25 A third object of the present invention is to provide vectors containing said genes.

A fourth object of the present invention is to provide hosts transformed with said vectors.

Another object of the present invention is to provide hosts which accumulate novel or uncommon carotenoids or which accumulate greater amounts of specific or total carotenoids.

30 Another object of the present invention is to provide hosts with inhibited and/or altered carotenoid production.

Another object of this invention is to secure the expression of eukaryotic carotenoid-related genes in a recombinant prokaryotic host.

Yet another object of the present invention is to provide a method for screening for eukaryotic and prokaryotic genes and cDNAs which encode enzymes involved in carotenoid biosynthesis and metabolism.

5 An additional object of the invention is to provide a method for manipulating carotenoid biosynthesis in photosynthetic organisms by inhibiting the synthesis of certain enzymatic products to cause accumulation of precursor compounds.

10 Another object of the invention is to provide modified lycopene  $\epsilon$ -cyclase, IPP isomerase and  $\beta$ -carotene hydroxylase.

These and other objects of the present invention have been realized by the present inventors as described below.

15 A subject of the present invention is an isolated and/or purified nucleic acid sequence which encodes for a protein having lycopene  $\epsilon$ -cyclase, IPP isomerase or  $\beta$ -carotene hydroxylase enzyme activity and having the amino acid sequence of SEQ ID NOS: 2, 4, 14-21 or 23-27.

20 The invention also includes vectors which comprise any of the nucleic acid sequences listed above, and host cells transformed with such vectors.

25 Another subject of the present invention is a method of producing or enhancing the production of a carotenoid in a host cell, comprising inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein having lycopene  $\epsilon$ -cyclase, IPP isomerase or  $\beta$ -carotene hydroxylase enzyme activity, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence to produce the protein.

30 Yet another subject of the present invention is a method of modifying the production of carotenoids in a host cell, the method comprising inserting into the host cell a vector comprising a heterologous nucleic acid sequence which produces an RNA and/or encodes for a protein which modifies lycopene  $\epsilon$ -cyclase, IPP isomerase or  $\beta$ -carotene hydroxylase enzyme activity, relative to an untransformed host cell, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence in the host cell to modify the production of the carotenoids in the host cell, relative to the untransformed host cell.

The present invention also includes a method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein having lycopene  $\epsilon$ -cyclase, IPP isomerase or  $\beta$ -carotene hydroxylase enzyme activity, the method comprising inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence.

5

Also included is a method of expressing, in a host cell, a heterologous nucleic acid sequence which encodes for a protein which modifies lycopene  $\epsilon$ -cyclase, IPP isomerase or  $\beta$ -carotene hydroxylase enzyme activity in the host cell, relative to an untransformed host cell, the method comprising inserting into the host cell a vector comprising the heterologous nucleic acid sequence, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence.

10

Another subject of the present invention is a method for screening for genes and cDNAs which encode enzymes involved in carotenoid biosynthesis and metabolism.

20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95

#### BRIEF DESCRIPTION OF THE DRAWINGS

A more complete appreciation of the invention and many of the attendant advantages thereof will be readily obtained as the same becomes better understood by reference to the following detailed description when considered in connection with the accompanying drawings, wherein:

25

Figure 1 is a schematic representation of the putative pathway of  $\beta$ -carotene biosynthesis in cyanobacteria, algae and plants. The enzymes catalyzing various steps are indicated at the left. Target sites of the bleaching herbicides NFZ and MPTA are also indicated at the left. Abbreviations: DMAPP, dimethylallyl pyrophosphate; FPP, farnesyl pyrophosphate; GGPP, geranylgeranyl pyrophosphate; GPP, geranyl pyrophosphate; IPP, isopentenyl pyrophosphate; LCY, lycopene cyclase; MVA, mevalonic acid; MPTA, 2-(4-methylphenoxy)triethylamine hydrochloride; NFZ, norflurazon; PDS, phytoene desaturase; PSY, phytoene synthase; ZDS,  $\zeta$ -carotene desaturase; PPPP, prephytoene pyrophosphate.

30

Figure 2 depicts possible routes of synthesis of cyclic carotenoids and common plant and algal xanthophylls (oxycarotenoids) from neurosporene. Demonstrated activities of the  $\beta$ - and  $\epsilon$ -cyclase enzymes of *A. thaliana* are indicated by bold arrows labelled with  $\beta$  or  $\epsilon$  respectively. A bar below the arrow leading to  $\epsilon$ -carotene indicates that the enzymatic

activity was examined but no product was detected. The steps marked by an arrow with a dotted line have not been specifically examined. Conventional numbering of the carbon atoms is given for neurosporene and  $\alpha$ -carotene. Inverted triangles (▼) mark positions of the double bonds introduced as a consequence of the desaturation reactions.

5 Figure 3 depicts the carotene endgroups which are found in plants.

Figure 4 is a DNA sequence and the predicted amino acid sequence of a lycopene  $\epsilon$ -cyclase cDNA isolated from *A. thaliana* (SEQ ID NOS: 1 and 2). These sequences were deposited under Genbank accession number U50738. This cDNA is incorporated into the plasmid pATeps.

10 Figure 5 is a DNA sequence encoding the  $\beta$ -carotene hydroxylase isolated from *A. thaliana* (SEQ ID NO: 3). This cDNA is incorporated into the plasmid pATOHB.

Figure 6 is an alignment of the predicted amino acid sequences of *A. thaliana*  $\beta$ -carotene hydroxylase (SEQ ID NO: 4) with those of the bacterial  $\beta$ -carotene hydroxylase enzymes from *Alicalgenes sp.* (SEQ ID NO: 5) (Genbank D58422), *Erwinia herbicola* Eho10 (SEQ ID NO.: 6) (GenBank M872280), *Erwinia uredovora* (SEQ ID NO.: 7) (GenBank D90087) and *Agrobacterium auranticum* (SEQ ID NO.: 8) (GenBank D58420). A consensus sequence is also shown. All five genes are identical where a capital letter appears in the consensus. A lowercase letter indicates that three of five, including *A. thaliana*, have the identical residue. TM; transmembrane.

20 Figure 7 is a DNA sequence of a cDNA encoding an IPP isomerase isolated from *A. thaliana* (SEQ ID NO: 9). This cDNA is incorporated into the plasmid pATDP5.

Figure 8 is a DNA sequence of a second cDNA encoding another IPP isomerase isolated from *A. thaliana* (SEQ ID NO: 10). This cDNA is incorporated into the plasmid pATDP7.

25 Figure 9 is a DNA sequence of a cDNA encoding an IPP isomerase isolated from *Haematococcus pluvialis* (SEQ ID NO: 11). This cDNA is incorporated into the plasmid pHp04.

Figure 10 is a DNA sequence of a second cDNA encoding another IPP isomerase isolated from *Haematococcus pluvialis* (SEQ ID NO: 12). This cDNA is incorporated into the plasmid pHp05.

30 Figure 11 is an alignment of the amino acid sequences predicted by IPP isomerase cDNAs isolated from *A. thaliana* (SEQ ID NO.: 16 and 18), *H. pluvialis* (SEQ ID NOS.: 14

and 15), *Clarkia breweri* (SEQ ID NO.: 17) (See, Blanc & Pichersky, *Plant Physiol.* (1995) 108:855; Genbank accession no. X82627) and *Saccharomyces cerevisiae* (SEQ ID NO.: 19) (Genbank accession no. J05090).

5 Figure 12 is a DNA sequence of the cDNA encoding an IPP isomerase isolated from *Tagetes erecta* (marigold; SEQ ID NO: 13). This cDNA is incorporated into the plasmid pPMDP1. xxx's denote a region not originally sequenced. Figure 21A shows the complete marigold sequence.

10 Figure 13 is an alignment of the consensus sequence of four plant  $\beta$ -cyclases (SEQ ID NO.: 20) with the *A. thaliana* lycopene  $\epsilon$ -cyclase (SEQ ID NO.: 21). A capital letter in the plant  $\beta$  consensus is used where all four  $\beta$ -cyclase genes predict the same amino acid residue in this position. A small letter indicates that an identical residue was found in three of the four. Dashes indicate that the amino acid residue was not conserved and dots in the sequence denote a gap. A consensus for the aligned sequences is given, in capital letters below the alignment, where the  $\beta$ - and  $\epsilon$ -cyclases have the same amino acid residue. Arrows indicate some of the conserved amino acids that will be used as junction sites for construction of chimeric cyclases with novel enzymatic activities. Several regions of interest including a sequence signature indicative of a dinucleotide-binding motif and two predicted transmembrane (TM) helical regions are indicated below the alignment and are underlined.

15 Figure 14 shows the nucleotide (SEQ ID NO:22) and amino acid sequences (SEQ ID NO:23) of the *Adonis palaestina* (pheasant's eye)  $\epsilon$ -cyclase cDNA #5.

20 Figure 15A shows the nucleotide (SEQ ID NO:24) and amino acid sequences (SEQ ID NO:25) of a potato  $\epsilon$ -cyclase cDNA. Figure 15B shows the amino acid sequence (SEQ ID NO:26) of a chimeric lettuce/potato lycopene  $\epsilon$ -cyclase. Amino acids in lower case are from the lettuce cDNA and those in upper case are from the potato cDNA. The product of this 25 chimeric cDNA has  $\epsilon$ -cyclase activity and converts lycopene to the monocyclic  $\delta$ -carotene.

Figure 16 shows a comparison between the amino acid sequences of the *Arabidopsis*  $\epsilon$ -cyclase (SEQ ID NO:27) and the potato  $\epsilon$ -cyclase (SEQ ID NO:25).

25 Figure 17A shows the nucleotide sequence of the *Adonis palaestina* Ipi1 (SEQ ID NO:28) and Figure 17B shows the nucleotide sequence of the *Adonis palaestina* Ipi2 (SEQ ID NO: 29).

Figure 18A shows the nucleotide sequence of the *Haematoccus pluvialis* Ipi1 (SEQ ID NO:11) and Figure 18B shows the nucleotide sequence of the *Haematoccus pluvialis* Ipi2 (SEQ ID NO:30).

5 Figure 19A shows the nucleotide sequence of the *Lactuca sativa* (*romaine lettuce*) Ipi1 (SEQ ID NO:31) and Figure 19B shows the nucleotide sequence of the *Lactuca sativa* Ipi2 (SEQ ID NO: 32).

Figure 20 shows the nucleotide sequence of the *Chlamydomonas reinhardtii* Ipi1 (SEQ ID NO:33).

10 Figure 21A shows the nucleotide sequence of the *Tagetes erecta* (marigold) Ipi1 (SEQ ID NO:34) and Figure 21B shows the nucleotide sequence of the *Oryza sativa* (rice) Ipi1 (SEQ ID NO:35).

15 Figure 22 shows a amino acid sequence alignment of various plant and green algal isopentenyl isomerase (IPI) (SEQ ID NOS:16, 36-45).

Figure 23 shows a comparison between *Adonis palaestina*  $\epsilon$ -cyclase cDNA #3 and cDNA #5 nucleotide sequences.

20 Figure 24 shows a comparison between *Adonis palaestina*  $\epsilon$ -cyclase cDNA #3 and cDNA #5 predicted amino acid sequences.

Figure 25 shows a sequence alignment of various plant  $\beta$ - and  $\epsilon$ -cyclases. Those sequences outlined in grey denote identical sequences among the  $\epsilon$ -cyclases. Those sequences outlined in black denote identical sequences among both the  $\beta$ - and  $\epsilon$ -cyclases.

25 Figure 26 shows a sequence alignment of the plant  $\epsilon$ -cyclases from Figure 25. Those sequences outlined in black denote identical sequences among the  $\epsilon$ -cyclases.

Figure 27 is a dendrogram or "tree" illustrating the degree of amino acid sequence similarity for various lycopene  $\beta$ - and  $\epsilon$ -cyclases.

30 Sub C2 Figure 28 shows a comparison between *Arabidopsis*  $\epsilon$ -cyclase and lettuce  $\epsilon$ -cyclase predicted amino acid sequences.

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention includes an isolated and/or purified nucleic acid sequence which encodes for a protein having lycopene  $\epsilon$ -cyclase, IPP isomerase or  $\beta$ -carotene hydroxylase enzyme activity and having the amino acid sequence of SEQ ID NOS: 2, 4, 14-21, 23 or 25-27. Nucleic acids encoding lycopene  $\epsilon$ -cyclase,  $\beta$ -carotene hydroxylase and IPP

isomerases have been isolated from several genetically distant sources.

The present inventors have isolated nucleic acids encoding the enzyme IPP isomerase, which catalyzes the reversible conversion of isopentenyl pyrophosphate (IPP) to dimethylallyl pyrophosphate (DMAPP). IPP isomerase cDNAs were isolated from the plants *A. thaliana*, *Tagetes erecta* (marigold), *Adonis palaestina* (pheasant's eye), *Lactuca sativa* (romaine lettuce) and from the green algae *H. pluvialis* and *Chlamydomonas reinhardtii*. Alignments of the amino acid sequences predicted by some of these cDNAs are shown in Figures 12 and 22. Plasmids containing some of these cDNAs were deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville MD 20852 on March 4, 1996 under ATCC accession numbers 98000 (pHP05 - *H. pluvialis*); 98001 (pMDP1 - marigold); 98002 (pATDP7 - *A. thaliana*) and 98004 (pHP04 - *H. pluvialis*).

The present inventors have also isolated nucleic acids encoding the enzyme  $\beta$ -carotene hydroxylase, which is responsible for hydroxylating the  $\beta$ -endgroup in carotenoids. The nucleic acid of the present invention is shown in SEQ ID NO: 3 and Figure 5. The full length cDNA product hydroxylates both end groups of  $\beta$ -carotene as do products of cDNAs which encode proteins truncated by up to 50 amino acids from the N-terminus. Products of genes which encode proteins truncated between about 60-110 amino acids from the N-terminus preferentially hydroxylate only one ring. A plasmid containing this gene was deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville MD 20852 on March 4, 1996 under ATCC accession number 98003 (pATOHB - *A. thaliana*).

The present inventors have also isolated nucleic acids encoding the enzyme lycopene  $\epsilon$ -cyclase, which is responsible for the formation of  $\epsilon$ -endgroups in carotenoids. The *A. thaliana*  $\epsilon$ -cyclase adds an  $\epsilon$  ring to only one end of the symmetrical lycopene while the related  $\beta$ -cyclase adds a ring at both ends. The *A. thaliana* cDNA of the present invention is shown in Figure 4 and SEQ ID NO: 1. A plasmid containing this gene was deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville MD 20852 on March 4, 1996 under ATCC accession number 98005 (pATeps - *A. thaliana*).

In addition, lycopene  $\epsilon$ -cyclases have been identified in lettuce and in *Adonis palaestina* (cDNA #5) which encode enzymes that convert lycopene to the bicyclic  $\epsilon$ -carotene ( $\epsilon,\epsilon$ -carotene). An additional cDNA from *Adonis palaestina* (cDNA #3) encodes a lycopene  $\epsilon$ -cyclase which converts lycopene into  $\delta$ -carotene ( $\epsilon,\psi$ -carotene) and differs from the lycopene  $\epsilon$ -cyclase which forms bicyclic  $\epsilon$ -carotene ( $\epsilon,\epsilon$ -carotene) by only 5 amino acids.

One or more of these amino acids may be modified by alteration of the nucleotide sequence in the #5 cDNA to obtain an enzyme which forms the bicyclic  $\epsilon,\epsilon$ -carotene. The sequences of the *Adonis palaestina* and *Arabidopsis thaliana*  $\epsilon$ -cyclases have about 70% nucleotide identity and about 72% amino acid identity.

5 Initial experiments by the inventors with chimeric genes indicated that the part of the  $\epsilon$ -cyclase which is responsible for adding 2  $\epsilon$  rings to form  $\epsilon,\epsilon$ -carotene is the carboxy terminal portion of the gene. The lettuce  $\epsilon$ -cyclase adds two  $\epsilon$  rings to form  $\epsilon,\epsilon$ -carotene. A DNA encoding a partial potato  $\epsilon$ -cyclase (missing its amino terminal portion), when combined with an amino terminal region from the lettuce  $\epsilon$ -cyclase gene, produces a 10 monocyclic  $\delta$ -carotene ( $\epsilon,\psi$ -carotene). With the discovery of the differences between the *Adonis palaestina* clone #3 and clone #5, the specific amino acids responsible for the addition of an extra  $\epsilon$  ring have been identified (Figure 24). Specifically, amino acid 55 is Thr in clone #3 and Ser in clone #5, amino acid 210 is Asn in clone #3 and Asp in clone #5, amino acid 231 is Asp in clone #3 and Glu in clone #5, amino acid 352 is Ile in clone #3 and Val in clone #5, and amino acid 524 is Lys in clone #3 and Arg in clone #5. It can be appreciated that these changes are quite conservative, as only one change, at amino acid 210, changes the charge of the protein.

15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Thus, it is clear that the nucleic acids of the invention encoding the enzymes as presently disclosed may be altered to increase a particularly desirable property of the enzyme, to change a property of the enzyme, or to diminish an undesirable property of the enzyme. Such modifications can be by deletion, substitution, or insertion of one or more amino acids, and can be performed by routine enzymatic manipulation of the nucleic acid encoding the enzyme (such as by restriction enzyme digestion, removal of nucleotides by mung bean nuclease or *Bal*31, insertion of nucleotides by Klenow fragment, and by religation of the ends), by site-directed mutagenesis, or may be accidental, such as by low fidelity PCR or those obtained through mutations in hosts that are producers of the enzymes. These 25 techniques as well as other suitable techniques are well known in the art.

30

Mutations can be made in the nucleic acids of the invention such that a particular codon is changed to a codon which codes for a different amino acid. Such a mutation is generally made by making the fewest nucleotide changes possible. A substitution mutation of this sort can be made to change an amino acid in the resulting protein in a non-conservative manner (i.e., by changing the codon from an amino acid belonging to a grouping

of amino acids having a particular size or characteristic to an amino acid belonging to another grouping) or in a conservative manner (i.e., by changing the codon from an amino acid belonging to a grouping of amino acids having a particular size or characteristic to an amino acid belonging to the same grouping). Such a conservative change generally leads to less change in the structure and function of the resulting protein. A non-conservative change is more likely to alter the structure, activity or function of the resulting protein. The present invention should be considered to include sequences containing conservative changes which do not significantly alter the activity or binding characteristics of the resulting protein.

5 The following is one example of various groupings of amino acids:

10 Amino acids with nonpolar R groups: Alanine, Valine, Leucine, Isoleucine, Proline, Phenylalanine, Tryptophan and Methionine.

Amino acids with uncharged polar R groups: Glycine, Serine, Threonine, Cysteine, Tyrosine, Asparagine and Glutamine.

Amino acids with charged polar R groups (negatively charged at Ph 6.0): Aspartic acid and Glutamic acid.

Basic amino acids (positively charged at pH 6.0): Lysine, Arginine and Histidine.

15 Another grouping may be those amino acids with phenyl groups: Phenylalanine, Tryptophan and Tyrosine.

Another grouping may be according to molecular weight (i.e., size of R groups).

20 Particularly preferred substitutions are:

- Lys for Arg and vice versa such that a positive charge may be maintained;
- Glu for Asp and vice versa such that a negative charge may be maintained;
- Ser for Thr such that a free -OH can be maintained; and
- Gln for Asn such that a free NH<sub>2</sub> can be maintained.

25 Amino acid substitutions may also be introduced to substitute an amino acid with a particularly preferable property. For example, a Cys may be introduced to provide a potential site for disulfide bridges with another Cys. A His may be introduced as a particularly "catalytic" site (i.e., His can act as an acid or base and is the most common amino acid in biochemical catalysis). Pro may be introduced because of its particularly planar structure, which induces  $\beta$ -turns in the protein's structure.

30 It is clear that certain modifications of SEQ ID NOS: 2, 4, 14-21, 23 or 25-27 can take place without destroying the activity of the enzyme. It is noted especially that truncated

versions of the nucleic acids of the invention are functional. For example, several amino acids (from 1 to about 120) can be deleted from the N-terminus of the lycopene  $\epsilon$ -cyclases of the invention, and a functional protein can still be produced. This fact is made especially clear from Figure 25, which shows a sequence alignment of several plant  $\epsilon$ -cyclases. As can be seen from Figure 25, there is an enormous amount of sequence disparity between amino acid sequences 2 to about 50-70 (depending on the particular sequence, since gaps are present). There is less, but also a substantial amount of, sequence dissimilarity between about 50-70 to about 90-120 (depending on the particular sequence). Thereafter, the sequences are fairly conserved, except for small pockets of dissimilarity between about 275-295 to about 285-305 (depending on the particular sequence), and between about 395-415 to about 410-430 (depending on the particular sequence).

The present inventors have found that the amount of the 5' region present in the nucleic acids of the invention can alter the activity of the enzyme. Instead of diminishing activity, truncating the 5' region of the nucleic acids of the invention may result in an enzyme with a different specificity. Thus, the present invention relates to nucleic acids and enzymes encoded thereby which are truncated to within 0-50, preferably 0-25, codons of the 5' initiation codon of their prokaryotic counterparts as determined by alignment maps as discussed below.

For example, when the cDNA encoding *A. thaliana*  $\beta$ -carotene hydroxylase was truncated, the resulting enzyme catalyzed the formation of  $\beta$ -cryptoxanthin as the major product and zeaxanthin as minor product; in contrast to its normal production of zeaxanthin.

The present invention is intended to include those nucleic acid and amino acid sequences in which substitutions, deletions, additions or other modifications have taken place, as compared to SEQ ID NOS: 2, 4, 14-21, 23 or 25-27, without destroying the activity of the enzyme. Preferably, the substitutions, deletions, additions or other modifications take place at the 5' end, or any other of those positions which already show dissimilarity between any of the presently disclosed amino acid sequences (see also Figure 25) or other amino acid sequences which are known in the art and which encode the same enzyme (i.e., lycopene  $\epsilon$ -cyclase, IPP isomerase or  $\beta$ -carotene hydroxylase).

In each case, nucleic acid and amino acid sequence similarity and identity is measured using sequence analysis software, for example, the Sequence Analysis, Gap, or BestFit software packages of the Genetics Computer Group (University of Wisconsin Biotechnology

Center, 1710 University Avenue, Madison, Wisconsin 53705), MEGAlign (DNAStar, Inc., 1228 S. Park St., Madison, Wisconsin 53715), or MacVector (Oxford Molecular Group, 2105 S. Bascom Avenue, Suite 200, Campbell, California 95008). Such software uses algorithms to match similar sequences by assigning degrees of identity to various substitutions, 5 deletions, and other modifications, and includes detailed instructions as to useful parameters, etc., such that those of routine skill in the art can easily compare sequence similarities and identities. An example of a useful algorithm in this regard is the algorithm of Needleman and Wunsch, which is used in the Gap program discussed above. This program finds the alignment of two complete sequences that maximizes the number of matches and minimizes 10 the number of gaps. Another useful algorithm is the algorithm of Smith and Waterman, which is used in the BestFit program discussed above. This program creates an optimal alignment of the best segment of similarity between two sequences. Optimal alignments are found by inserting gaps to maximize the number of matches using the local homology algorithm of Smith and Waterman.

15 Conservative (i.e. similar) substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid, glutamic acid, asparagine and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine. Substitutions may also be made on the basis of conserved hydrophobicity or hydrophilicity (see Kyte and Doolittle, *J. Mol. Biol.* 157: 105-132 (1982)), or on the basis of the ability to assume similar polypeptide secondary structure (see Chou and Fasman, *Adv. Enzymol.* 47: 45-148 (1978)).

20 If comparison is made between nucleotide sequences, preferably the length of comparison sequences is at least 50 nucleotides, more preferably at least 60 nucleotides, at least 75 nucleotides or at least 100 nucleotides. It is most preferred if comparison is made between the nucleic acid sequences encoding the enzyme coding regions necessary for enzyme activity. If comparison is made between amino acid sequences, preferably the length of comparison is at least 20 amino acids, more preferably at least 30 amino acids, at least 40 amino acids or at least 50 amino acids. It is most preferred if comparison is made between 25 the amino acid sequences in the enzyme coding regions necessary for enzyme activity.

30 It should be appreciated that also within the scope of the present invention are nucleic acid sequences encoding lycopene  $\epsilon$ -cyclases, IPP isomerases and  $\beta$ -carotene hydroxylases

which code for enzymes having the same amino acid sequence as SEQ ID NOS: 2, 4, 14-21, 23 or 25-27, but which are degenerate to the nucleic acids specifically disclosed herein.

5 The amino acid residues described herein are preferred to be in the "L" isomeric form. However, residues in the "D" isomeric form can be substituted for any L-amino acid residue, as long as the desired functional property of immunoglobulin-binding is retained by the polypeptide.

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook et al, 10 "Molecular Cloning: A Laboratory Manual" (1989); "Current Protocols in Molecular Biology" Volumes I-III [Ausubel, R. M., ed. (1994)]; "Cell Biology: A Laboratory Handbook" Volumes I-III [J. E. Celis, ed. (1994)]; "Current Protocols in Immunology" Volumes I-III [Coligan, J. E., ed. (1994)]; "Oligonucleotide Synthesis" (M.J. Gait ed. 1984); "Nucleic Acid Hybridization" [B.D. Hames & S.J. Higgins eds. (1985)]; "Transcription And Translation" [B.D. Hames & S.J. Higgins, eds. (1984)]; "Animal Cell Culture" [R.I. Freshney, ed. (1986)]; "Immobilized Cells And Enzymes" [IRL Press, (1986)]; B. Perbal, "A Practical Guide To Molecular Cloning" (1984).

The present invention also includes vectors. Suitable vectors according to the present invention comprise a nucleic acid of the invention encoding an enzyme involved in carotenoid biosynthesis or metabolism and a suitable promoter for the host, and can be constructed using techniques well known in the art (for example Sambrook et al., Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989; Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing and Wiley Interscience, New York, 1991). Suitable vectors for eukaryotic expression in plants are described in Frey et al., Plant J. (1995) 8(5):693 and Misawa et al, 1994a; incorporated herein by reference. Suitable vectors for prokaryotic expression include pACYC184, pUC119, and pBR322 (available from New England BioLabs, Beverly, MA) and pTrcHis (Invitrogen) and pET28 (Novagen) and derivatives thereof. The vectors of the present invention can additionally contain regulatory elements such as promoters, repressors, selectable markers such as antibiotic resistance genes, etc.

30 The nucleic acids encoding the carotenoid enzymes as described above, when cloned into a suitable expression vector, can be used to overexpress these enzymes in a plant

5

expression system or to inhibit the expression of these enzymes. For example, a vector containing the gene encoding lycopene  $\epsilon$ -cyclase can be used to increase the amount of  $\alpha$ -carotene and carotenoids derived from  $\alpha$ -carotene (such as lutein and  $\alpha$ -cryptoxanthin) in an organism and thereby alter the nutritional value, pharmacology and visual appearance value of the organism.

10

Therefore, the present invention includes a method of producing or enhancing the production of a carotenoid in a host cell, relative to an untransformed host cell, the method comprising inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein having lycopene  $\epsilon$ -cyclase, IPP isomerase or  $\beta$ -carotene hydroxylase enzyme activity, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence to produce the protein.

25

30

The present invention also includes a method of modifying the production of carotenoids in a host cell, the method comprising inserting into the host cell a vector comprising a heterologous nucleic acid sequence which produces an RNA and/or encodes for a protein which modifies lycopene  $\epsilon$ -cyclase, IPP isomerase or  $\beta$ -carotene hydroxylase enzyme activity, relative to an untransformed host cell, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and expressing the heterologous nucleic acid sequence in the host cell to modify the production of the carotenoids in the host cell, relative to the untransformed host cell.

The term "modifying the production" means that the amount of carotenoids produced in the host cell can be enhanced, reduced, or left the same, as compared to the untransformed host cell. In accordance with one embodiment of the present invention, the make-up of the carotenoids (i.e., the specific carotenoids produced) is changed *vis a vis* each other, and this change in make-up may result in either a net gain, net loss, or no net change in the total amount of carotenoids produced in the cell. In accordance with another embodiment of the present invention, the production or the biochemical activity of the carotenoids (or the enzymes which catalyze their formation) is enhanced by the insertion of an enzyme-encoding nucleic acid of the invention. In yet another embodiment of the invention, the production or the biochemical activity of the carotenoids (or the enzymes which catalyze their formation) may be reduced or inhibited by a number of different approaches available to those skilled in the art, including but not limited to such methodologies or approaches as anti-sense (e.g.,

Gray et al (1992) *Plant Mol. Biol.* 19:69-87), ribozymes (e.g., Wegener et al (1994) *Mol. Gen. Genet.* 245:465-470), co-suppression (e.g., Fray and Grierson (1993) *Plant Mol. Biol.* 22:589-602), targeted disruption of the gene (e.g., Schaefer et al. (1997) *Plant J.* 11:1195-1206), intracellular antibodies (e.g., Rondon and Marasco (1997) *Ann. Rev. Microbiol.* 51:257-283) or whatever other approaches rely on the knowledge or availability of the nucleic acid or amino acid sequences of the invention and/or portions thereof, to thereby reduce accumulation of carotenoids with  $\epsilon$  rings and compounds derived from them (for  $\epsilon$ -cyclase inhibition), or carotenoids with hydroxylated  $\beta$  rings and compounds derived from them (for  $\beta$ -hydroxylase inhibition), or, in the case of IPP isomerase, accumulation of any isoprenoid compound.

10 Preferably, at least a portion of the nucleic acid sequences used in the methods, vectors and host cells of the invention codes for an enzyme having an amino acid sequence which is at least 85% identical, preferably at least 90%, at least 95% or completely identical to SEQ ID NOS: 2, 4, 14-21, 23 or 25-27. Sequence identity is determined as noted above. Preferably, sequence additions, deletions or other modifications are made as indicated above, so as to not affect the function of the particular enzyme.

15 In a preferred embodiment, vectors are manufactured which contain a DNA encoding a eukaryotic IPP isomerase upstream of a DNA encoding a second eukaryotic carotenoid enzyme. The inventors have discovered that inclusion of an IPP isomerase gene increases the supply of substrate for the carotenoid pathway; thereby enhancing the production of carotenoid endproducts, as compared to a host cell which is not transformed with such a vector. This is apparent from the much deeper pigmentation in carotenoid-accumulating colonies of *E. coli* which also contain one of the aforementioned IPP isomerase genes when compared to colonies that lack this additional IPP isomerase gene. Similarly, a vector comprising an IPP isomerase gene can be used to enhance production of any secondary metabolite of dimethylallyl pyrophosphate and/or isopentenyl pyrophosphate (such as isoprenoids, steroids, carotenoids, etc.). The term "isoprenoid" is intended to mean any member of the class of naturally occurring compounds whose carbon skeletons are composed, in part or entirely, of isopentyl C<sub>5</sub> units. Preferably, the carbon skeleton is of an essential oil, a fragrance, a rubber, a carotenoid, or a therapeutic compound, such as paclitaxel.

20 25 30 A vector containing the cDNA encoding a lycopene  $\epsilon$ -cyclase of the invention, preferably the lettuce lycopene  $\epsilon$ -cyclase or *Adonis*  $\epsilon$ -cyclase #5, can be used to increase the

amount of bicyclic  $\epsilon$ -carotene in an organism and thereby alter the nutritional value, pharmacology and visual appearance value of the organism. In addition, the transformed organism can be used in the formulation of therapeutic agents, for example in the treatment of cancer (see Mayne et al (1996) FASEB J. 10:690-701; Tsushima et al (1995) Biol. Pharm. Bull. 18:227-233).

5

An antisense strand of a nucleic acid of the invention can be inserted into a vector. For example, the lycopene  $\epsilon$ -cyclase gene can be inserted into a vector and incorporated into the genomic DNA of a host, thereby inhibiting the synthesis of  $\epsilon, \beta$ -carotenoids (lutein and  $\alpha$ -carotene) and enhancing the synthesis of  $\beta, \beta$ -carotenoids (zeaxanthin and  $\beta$ -carotene).

10 The present invention also relates to novel enzymes which are encoded by the amino acid sequences of the invention, or portions thereof.

The present invention also relates to novel enzymes which can transform known carotenoids into novel or uncommon products. Currently  $\epsilon$ -carotene (see Figure 2) and  $\gamma$ -carotene are commonly produced only in minor amounts. As described below, an enzyme can be produced which transforms lycopene to  $\gamma$ -carotene and lycopene to  $\epsilon$ -carotene. With these products in hand, bulk synthesis of other carotenoids derived from them are possible. For example,  $\epsilon$ -carotene can be hydroxylated to form lactucaxanthin, an isomer of lutein (one  $\epsilon$  and one  $\beta$  ring) and zeaxanthin (two  $\beta$  rings) where both endgroups are, instead,  $\epsilon$  rings.

15 In addition to novel enzymes produced by truncating the 5' region of known enzymes, as discussed above, novel enzymes which can participate in the formation of unusual carotenoids can be formed by replacing portions of one gene with an analogous sequence from a structurally related gene. For example,  $\beta$ -cyclase and  $\epsilon$ -cyclase are structurally related (see Figure 13). By replacing a portion of  $\beta$ -lycopene cyclase with the analogous portion of  $\epsilon$ -cyclase, an enzyme which produces  $\gamma$ -carotene will be produced (one  $\beta$  endgroup). Further, by replacing a portion of the lycopene  $\epsilon$ -cyclase with the analogous portion of  $\beta$ -cyclase, an enzyme which produces  $\epsilon$ -carotene will be produced (with some exceptions, such as the lettuce  $\epsilon$ -cyclase, plant  $\epsilon$ -cyclases normally produce a compound with one  $\epsilon$ -endgroup,  $\delta$ -carotene). Similarly,  $\beta$ -hydroxylase could be modified to produce enzymes of novel function by creation of hybrids with  $\epsilon$ -hydroxylase.

25

20 Host systems according to the present invention can comprise any organism that already produces carotenoids or which has been genetically modified to produce carotenoids.

30

The IPP isomerase genes are more broadly applicable for enhancing production of any product dependent on DMAPP and/or IPP as a precursor.

Organisms which already produce carotenoids include plants, algae, some yeasts, fungi and cyanobacteria and other photosynthetic bacteria. Transformation of these hosts with vectors according to the present invention can be done using standard techniques such as those described in Misawa et al., (1990) *supra*; Hundle et al., (1993) *supra*; Hundle et al., (1991) *supra*; Misawa et al., (1991) *supra*; Sandmann et al., *supra*; and Schnurr et al., *supra*.

Transgenic organisms can be constructed which include the nucleic acid sequences of the present invention (Bird et al, 1991; Bramley et al, 1992; Misawa et al, 1994a; Misawa et al, 1994b; Cunningham et al, 1993). The incorporation of these sequences can allow the controlling of carotenoid biosynthesis, content, or composition in the host cell. These transgenic systems can be constructed to incorporate sequences which allow for the overexpression of the nucleic acids of the present invention. Transgenic systems can also be constructed containing antisense expression of the nucleic acid sequences of the present invention. Such antisense expression would result in the accumulation of the substrates of the substrates of the enzyme encoded by the sense strand.

A method for screening for eukaryotic genes which encode enzymes involved in carotenoid biosynthesis comprises transforming a prokaryotic host with a nucleic acid which may contain a eukaryotic or prokaryotic carotenoid biosynthetic gene; culturing said transformed host to obtain colonies; and screening for colonies exhibiting a different color than colonies of the untransformed host.

Suitable hosts include *E. coli*, cyanobacteria such as *Synechococcus* and *Synechocystis*, alga and plant cells. *E. coli* are preferred.

In a preferred embodiment, the above "color complementation" screening protocol can be enhanced by using mutants which are either (1) deficient in at least one carotenoid biosynthetic gene or (2) overexpress at least one carotenoid biosynthetic gene. In either case, such mutants will accumulate carotenoid precursors.

Prokaryotic and eukaryotic DNA or cDNA libraries can be screened in total for the presence of genes of carotenoid biosynthesis, metabolism and degradation. Preferred organisms to be screened include photosynthetic organisms.

5 *E. coli* can be transformed with these eukaryotic cDNA libraries using conventional methods such as those described in Sambrook et al, 1989 and according to protocols described by the vendors of the cloning vectors.

10 For example, the cDNA libraries in bacteriophage vectors such as lambdaZAP (Stratagene) or lambda ZIPLOX (Gibco BRL) can be excised en masse and used to transform *E.coli*.

15 Transformed *E. coli* can be cultured using conventional techniques. The culture broth preferably contains antibiotics to select and maintain plasmids. Suitable antibiotics include penicillin, ampicillin, chloramphenicol, etc. Culturing is typically conducted at 15-40°C, preferably at room temperature or slightly above (18-28°C), for 12 hours to 7 days.

20 Cultures are plated and the plates are screened visually for colonies with a different color than the colonies of the host *E. coli* transformed with the empty plasmid cloning vector. For example, *E. coli* transformed with the plasmid, pAC-BETA (described below), produce yellow colonies that accumulate  $\beta$ -carotene. After transformation with a cDNA library, colonies which contain a different hue than those formed by *E. coli*/pAC-BETA would be expected to contain enzymes which modify the structure or accumulation of  $\beta$ -carotene. Similar *E. coli* strains can be engineered which accumulate earlier products in carotenoid biosynthesis, such as lycopene,  $\gamma$ -carotene, etc.

25 Having generally described this invention, a further understanding can be obtained by reference to certain specific examples which are provided herein for purposes of illustration only and are not intended to be limiting unless otherwise specified.

## EXAMPLE

### I. Isolation of $\beta$ -carotene hydroxylase

#### Plasmid Construction

30 An 8.6kb BglII fragment containing the carotenoid biosynthetic genes of *Erwinia herbicola* was first cloned in the BamHI site of plasmid vector pACYC184 (chloramphenicol resistant), and then a 1.1kb BamHI fragment containing the *E. herbicola*  $\beta$ -carotene hydroxylase (*CrtZ*) was deleted. *E.coli* strains containing the resulting plasmid, pAC-BETA, accumulate  $\beta$ -carotene and form yellow colonies (Cunningham et al., 1994).

35 A full length cDNA encoding IPP isomerase of *Haematococcus pluvialis* (HP04) was first excised with *BamH*I and *Kpn*I from pBluescript SK-, and then ligated into the

corresponding sites of the pTrcHisA vector with high-level expression from the *trc* promoter (Invitrogen, Inc.). A fragment containing the IPP isomerase and *trc* promoter was subsequently excised with *EcoRV* and *KpnI*, treated with the Klenow fragment of DNA polymerase to produce blunt ends, and ligated in the Klenow-treated *HindIII* site of pAC-BETA. 5 *E.coli* cells transformed with this new plasmid pAC-BETA-04 form orange colonies on LB plates (vs. yellow for those containing pAC-BETA) and cultures accumulate substantially more  $\beta$ -carotene (ca. two fold) than those that contain pAC-BETA.

#### Screening of an *Arabidopsis* cDNA Library

10 Several  $\lambda$  cDNA expression libraries of *Arabidopsis* were obtained from the *Arabidopsis* Biological Resource Center (Ohio State University, Columbus, OH) (Kieber et al., 1993). The  $\lambda$  cDNA libraries were excised *in vivo* using Stratagene's ExAssist SOLR system to produce a phagemid cDNA library wherein each phagemid contained also a gene conferring resistance to the antibiotic ampicillin.

15 *E.coli* strain DH10BZIP was chosen as the host cell for the screening and pigment production, although we have also used TOP10F' and XL1-Blue for this purpose. DH10B cells were transformed with plasmid pAC-BETA-04 and were plated on LB agar plates containing chloramphenicol at 50  $\mu$ g/ml (from United States Biochemical Corporation). The phagemid *Arabidopsis* cDNA library was then introduced into DH10B cells already containing pAC-BETA-04. Transformed cells containing both pAC-BETA-04 and 20 *Arabidopsis* cDNA library phagemids were selected on chloramphenicol plus ampicillin (150  $\mu$ g/ml) agar plates. Maximum color development occurred after 3 to 7 days incubation at room temperature, and the rare bright yellow colonies were selected from a background of many thousands of orange colonies on each agar plate. Selected colonies were inoculated into 3 ml liquid LB medium containing ampicillin and chloramphenicol, and cultures were 25 incubated at room temperature for 1-2 days, with shaking. Cells were then harvested by centrifugation and extracted with acetone in microfuge tubes. After centrifugation, the pigmented extract was spotted onto silica gel thin-layer chromatography (TLC) plates, and developed with a hexane:ether (1:1, by volume) mobile phases.  $\beta$ -carotene hydroxylase-encoding cDNAs were identified based on the appearance of a yellow pigment that co-migrated with zeaxanthin on the TLC plates.

### Subcloning and Sequencing

The plasmid containing the  $\beta$ -carotene hydroxylase cDNA was recovered and analyzed by standard procedures (Sambrook et al., 1989). The *Arabidopsis*  $\beta$ -carotene hydroxylase was sequenced completely on both strands on an automatic sequencer (Applied Biosystems, Model 373A, Version 2.0.1S). The cDNA insert of 0.95kb also was excised and ligated into the a pTrcHis vector. A *Bgl*II restriction site within the cDNA was used to remove that portion of the cDNA that encodes the predicted polypeptide N terminal sequence region that is not also found in bacterial  $\beta$ -carotene hydroxylases (Figure 6). A *Bgl*III-*Xho*I fragment was directionally cloned in *Bam*HI-*Xho*I digested TrcHis vectors.

### 10 **Pigment Analysis**

A single colony was used to inoculate 50 ml of LB containing ampicillin and chloramphenicol in a 250-ml flask. Cultures were incubated at 28°C for 36 hours with gentle shaking, and then harvested at 5000 rpm in an SS-34 rotor. The cells were washed once with distilled H<sub>2</sub>O and resuspended with 0.5 ml of water. The extraction procedures and HPLC were essentially as described previously (Cunningham et al, 1994).

## II. Isolation and biochemical analysis of an *Arabidopsis* lycopene $\epsilon$ -cyclase

### Plasmid Construction

Construction of plasmids pAC-LYC, pAC-NEUR, and pAC-ZETA is described in Cunningham et al., (1994). In brief, the appropriate carotenoid biosynthetic genes from *Erwinia herbicola*, *Rhodobacter capsulatus*, and *Synechococcus* sp. strain PCC7942 were cloned in the plasmid vector pACYC184 (New England BioLabs, Beverly, MA). Cultures of *E. coli* containing the plasmids pAC-ZETA, pAC-NEUR, and pAC-LYC, accumulate  $\zeta$ -carotene, neurosporene, and lycopene, respectively. The plasmid pAC-ZETA was constructed as follows: an 8.6-kb *Bgl*II fragment containing the carotenoid biosynthetic genes of *E. herbicola* (GenBank M87280; Hundle et al., 1991) was obtained after partial digestion of plasmid pPL376 (Perry et al., 1986; Tuveson et al., 1986) and cloned in the *Bam*HI site of pACYC184 to give the plasmid pAC-EHER. Deletion of adjacent 0.8- and 1.1-kb *Bam*HI-*Bam*HI fragments (deletion Z in Cunningham et al., 1994), and of a 1.1 kB *Sall*-*Sall* fragment (deletion X) served to remove most of the coding regions for the *E. herbicola*  $\beta$ -carotene hydroxylase (*crtZ* gene) and zeaxanthin glucosyltransferase (*crtX* gene), respectively. The

resulting plasmid, pAC-BETA, retains functional genes for geranylgeranyl pyrophosphate synthase (crtE), phytoene synthase (crtB), phytoene desaturase (crtI), and lycopene cyclase (crtY). Cells of *E. coli* containing this plasmid form yellow colonies and accumulate  $\beta$ -carotene. A plasmid containing both the lycopene  $\epsilon$ - and  $\beta$ -cyclase cDNAs of *A. thaliana* was constructed by excising the  $\epsilon$ -cyclase in clone y2 as a PvuI-PvuII fragment and ligating this piece in the SnaBI site of a plasmid (pSPORT 1 from GIBCO-BRL) that already contained the  $\beta$ -cyclase (Cunningham et al., 1996).

### Organisms and Growth Conditions

*E. coli* strains TOP10 and TOP10 F' (obtained from Invitrogen Corporation, San Diego, CA) and XL1-Blue (Stratagene) were grown in Luria-Bertani (LB) medium (Sambrook et al., 1989) at 37°C in darkness on a platform shaker at 225 cycles per min. Media components were from Difco (yeast extract and tryptone) or Sigma (NaCl). Ampicillin at 150  $\mu$ g/mL and/or chloramphenicol at 50  $\mu$ g/mL (both from United States Biochemical Corporation) were used, as appropriate, for selection and maintenance of plasmids.

### Mass Excision and Color Complementation Screening of an *A. thaliana* cDNA Library

A size-fractionated 1-2 kB cDNA library of *A. thaliana* in lambda ZAPII (Kieber et al., 1993) was obtained from the Arabidopsis Biological Resource Center at The Ohio State University (stock number CD4-14). Other size fractionated libraries were also obtained (stock numbers CD4-13, CD4-15, and CD4-16). An aliquot of each library was treated to cause a mass excision of the cDNAs and thereby produce a phagemid library according to the instructions provided by the supplier of the cloning vector (Stratagene; *E. coli* strain XL1-Blue and the helper phage R408 were used). The titre of the excised phagemid was determined and the library was introduced into a lycopene-accumulating strain of *E. coli* TOP10 F' (this strain contained the plasmid pAC-LYC) by incubation of the phagemid with the *E. coli* cells for 15 min at 37°C. Cells had been grown overnight at 30°C in LB medium supplemented with 2% (w/v) maltose and 10 mM MgSO<sub>4</sub> (final concentration), and harvested in 1.5 ml microfuge tubes at a setting of 3 on an Eppendorf microfuge (5415C) for 10 min. The pellets were resuspended in 10 mM MgSO<sub>4</sub> to a volume equal to one-half that of the

initial culture volume. Transformants were spread on large (150 mm diameter) LB agar petri plates containing antibiotics to provide for selection of cDNA clones (ampicillin) and maintenance of pAC-LYC (chloramphenicol). Approximately 10,000 colony forming units were spread on each plate. Petri plates were incubated at 37°C for 16 hr and then at room 5 temperature for 2 to 7 days to allow maximum color development. Plates were screened visually with the aid of an illuminated 3x magnifier and a low power stage-dissecting microscope for the rare, pale pinkish-yellow to deep-yellow colonies that could be observed in the background of pink colonies. A colony color of yellow or pinkish-yellow was taken as presumptive evidence of a cyclization activity. These yellow colonies were collected with 10 sterile toothpicks and used to inoculate 3ml of LB medium in culture tubes with overnight growth at 37°C and shaking at 225 cycles/min. Cultures were split into two aliquots in microfuge tubes and harvested by centrifugation at a setting of 5 in an Eppendorf 5415C microfuge. After discarding the liquid, one pellet was frozen for later purification of plasmid DNA. To the second pellet was added 1.5 ml EtOH, and the pellet was resuspended by vortex mixing, and extraction was allowed to proceed in the dark for 15-30 min with occasional remixing. Insoluble materials were pelleted by centrifugation at maximum speed for 10 min in a microfuge. Absorption spectra of the supernatant fluids were recorded from 350-550 nm with a Perkin Elmer lambda six spectrophotometer.

#### Analysis of isolated clones

25

Eight of the yellow colonies contained  $\beta$ -carotene indicating that a single gene product catalyzes both cyclizations required to form the two  $\beta$  endgroups of the symmetrical  $\beta$ -carotene from the symmetrical precursor lycopene. One of the yellow colonies contained a pigment with the spectrum characteristic of  $\delta$ -carotene, a monocyclic carotenoid with a single  $\epsilon$  endgroup. Unlike the  $\beta$  cyclase, this  $\epsilon$ -cyclase appears unable to carry out a second cyclization at the other end of the molecule.

The observation that  $\epsilon$ -cyclase is unable to form two cyclic  $\epsilon$ -endgroups (e.g. the bicyclic  $\epsilon$ -carotene) illuminates the mechanism by which plants can coordinate and control the flow of substrate into carotenoids derived from  $\beta$ -carotene versus those derived from  $\alpha$ -carotene and also can prevent the formation of carotenoids with two  $\epsilon$  endgroups.

30

The availability of the *A. thaliana* gene encoding the  $\epsilon$ -cyclase enables the directed manipulation of plant and algal species for modification of carotenoid content and

composition. Through inactivation of the  $\epsilon$ -cyclase, whether at the gene level by deletion of the gene or by insertional inactivation or by reduction of the amount of enzyme formed (by such as antisense technology), one may increase the formation of  $\beta$ -carotene and other pigments derived from it. Since vitamin A is derived only from carotenoids with  $\beta$  endgroups, an enhancement of the production of  $\beta$ -carotene versus  $\alpha$ -carotene may enhance nutritional value of crop plants. Reduction of carotenoids with  $\epsilon$ -endgroups may also be of value in modifying the color properties of crop plants and specific tissues of these plants. Alternatively, where production of  $\alpha$ -carotene, or pigments such as lutein that are derived from  $\alpha$ -carotene, is desirable, whether for the color properties, nutritional value or other reason, one may overexpress the  $\epsilon$ -cyclase or express it in specific tissues. Wherever agronomic value of a crop is related to pigmentation provided by carotenoid pigments the directed manipulation of expression of the  $\epsilon$ -cyclase gene and/or production of the enzyme may be of commercial value.

The predicted amino acid sequence of the *A. thaliana*  $\epsilon$ -cyclase enzyme was determined. A comparison of the amino acid sequences of the  $\beta$ - and  $\epsilon$ -cyclase enzymes of *Arabidopsis thaliana* (Fig. 13) as predicted by the DNA sequence of the respective cDNAs (Fig. 4 for the  $\epsilon$ -cyclase cDNA sequence), indicates that these two enzymes have many regions of sequence similarity, but they are only about 37% identical overall at the amino acid level. The degree of sequence identity at the DNA base level, only about 50%, is sufficiently low such that we and others have been unable to detect this gene by hybridization using the  $\beta$  cyclase as a probe in DNA gel blot experiments.

#### REFERENCES

Each reference cited in this application and/or listed below is hereby incorporated by reference.

25                   Bird et al, 1991 Biotechnology 9, 635-639.  
                      Bishop et al., (1995) FEBS Lett. 367, 158-162.  
                      Bramley, P.M. (1985) Adv. Lipid Res. 21, 243-279.  
                      Bramley, P.M. (1992) Plant J. 2, 343-349.  
                      Britton, G. (1988). Biosynthesis of carotenoids. In Plant Pigments, T.W. Goodwin, ed.  
30                   (London: Academic Press), pp. 133-182.

5

Britton, G. (1979) *Z. Naturforsch. Section C Biosci.* 34, 979-985.

Britton, G. (1995) UV/Visible spectroscopy. In *Carotenoids*, Vol. IB: Spectroscopy, G. Britton, S. Liaaen-Jensen, H.P. Pfander, eds. (Basel: Birkhauser Verlag), pp. 13-62.

Bouvier et al., (1994) *Plant J.* 6, 45-54.

Cunningham et al., (1985) *Photochem. Photobiol.* 42: 295-307.

Cunningham et al., (1993) *FEBS Lett.* 328, 130-138.

Cunningham et al., (1994) *Plant Cell* 6, 1107-1121.

Cunningham et al., (1996) *Plant Cell* 8, 1613-1626.

Davies, B.H. (1976). Carotenoids. In *Chemistry and Biochemistry of Plant Pigments*, Vol. 2, T.W. Goodwin, ed (New York: Academic Press), pp. 38-165.

Del Sal et al., (1988). *Nucl. Acids Res.* 16, 9878.

Demmig-Adams & Adams, (1992) *Ann. Rev. Plant Physiol. Mol. Biol.* 43, 599-626.

Enzell & Back, (1995) Mass spectrometry. In *Carotenoids*, Vol. IB: Spectroscopy, G. Britton, S. Liaaen-Jensen, H.P. Pfander, eds. (Basel: Birkhauser Verlag), pp. 261-320.

Frank & Cogdell (1993) Photochemistry and function of carotenoids in photosynthesis. In *Carotenoids in Photosynthesis*. A. Young and G. Britton, eds. (London: Chapman and Hall). pp. 253-326.

Goodwin, T.W. (1980). *The Biochemistry of the Carotenoids*. 2nd ed, Vol. 1 (London: Chapman and Hall).

Horvath et al., (1972) *Phytochem.* 11, 183-187.

Hugueney et al., (1995) *Plant J.* 8, 417-424.

Hundle et al., (1991) *Photochem. Photobiol.* 54, 89-93.

Jensen & Jensen, (1971) *Methods Enzymol.* 23, 586-602.

Kargl & Quackenbush, (1960) *Archives Biochem. Biophys.* 88, 59-63.

25 Kargl et al., (1960) *Proc. Am. Hort. Soc.* 75, 574-578.

Kieber et al., (1993) *Cell* 72, 427-441.

Koyama, Y. (1991) *J. Photochem. Photobiol., B*, 9, 265-80.

Krinsky, N.I. (1987) Medical uses of carotenoids. In *Carotenoids*, N.I. Krinsky, M.M. Mathews-Roth, and R.F. Taylor, eds. (New York: Plenum), pp. 195-206.

30 Kyte & Doolittle, (1982) *J. Mol. Biol.* 157, 105-132.

LaRossa & Schloss, (1984) *J. Biol. Chem.* 259, 8753-8757.

Misawa et al., (1994a) *Plant J.* 6, 481-489.

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95

5

Misawa et al., (1994b) *J. Biochem. Tokyo*, 116, 980-985.

Norris et al., (1995) *Plant Cell* 7, 2139-2149.

Pecker et al., (1996) Submitted to *Plant Mol. Biol.*

Perry et al., (1986) *J. Bacteriol.* 168, 607-612.

Persson & Argos, (1994) *J. Mol. Biol.* 237, 182-192.

Plumley & Schmidt, (1987) *Proc. Nat. Acad. Sci. USA* 83, 146-150.

Plumley & Schmidt, (1995) *Plant Cell* 7, 689-704.

Rossmann et al., (1974) *Nature* 250, 194-199.

Rock & Zeevaart (1991) *Proc. Nat. Acad. Sci. USA* 88, 7496-7499.

Rost et al., (1995) *Protein Science* 4, 521-533.

Sambrook et al., (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Sancar, A. (1994) *Biochemistry* 33, 2-9.

Sander & Schneider, (1991) *Proteins* 9, 56-68.

Sandmann, G. (1994) *Eur. J. Biochem.* 223, 7-24.

Scolnik & Bartley, (1995) *Plant Physiol.* 108, 1342.

Siefermann-Harms, D. (1987) *Physiol. Plant.* 69, 561-568.

Spurgeon & Porter, (1980). Biosynthesis of carotenoids. In *Biochemistry of Isoprenoid Compounds*, J.W. Porter, and S.L. Spurgeon, eds. (New York: Wiley), pp. 1-122.

Tomes, M.L. (1963) *Bot. Gaz.* 124, 180-185.

Tomes, M.L. (1967) *Genetics* 56, 227-232.

Tuveson et al., (1986) *J. Bacteriol.* 170, 4675-4680.

Van Beeumen et al., (1991) *J. Biol. Chem.* 266, 12921-12931.

Weedon & Moss, (1995) Structure and Nomenclature. In *Carotenoids, Vol. IB: Spectroscopy*, G. Britton, S. Liaaen-Jensen, H.P. Pfander, eds. (Basel: Birkhauser Verlag), pp. 27-70.

Wierenga et al., (1986) *J. Mol. Biol.* 187, 101-107.

Zechmeister, L. (1962) *Cis-Trans Isomeric Carotenoids, Vitamins A and Arylpolyenes*. Springer-Verlag, Vienna.

Having now fully described the invention, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the invention as set forth herein.

15  
20  
25